JNATIA Volume 3, Nomor 3, Mei 2025 p-ISSN: 2986-3929
Jurnal Nasional Teknologi Informasi dan Aplikasinya e-ISSN: 3032-1948

Perbandingan Performa Ensemble Classifier dan Model Klasifikasi Tunggal dalam Clickbait Detection

Ilham Arsy Dwi Atmojo[a1], I Gede Arta Wibawa[a2]

Program Studi Informatika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Udayana
Jalan Raya Kampus UNUD, Bukit Jimbaran, Kuta Selatan, Badung, Bali, Indonesia
[1]ilhamarsy16@gmail.com
[2]gede.arta@unud.ac.id

Abstract

The rapid advancement of the digital era has led to a significant increase in online news content, accompanied by a growing issue of clickbait—sensational headlines designed to attract readers. This study aims to develop a clickbait detection system to classify Indonesian news headlines as clickbait or non-clickbait. Building upon previous research, this work explores the performance of ensemble classifiers compared to single classifiers such as Multinomial Naïve Bayes, Support Vector Machine (SVM), and Random Forest. Using a dataset of 4,500 headlines with 51.6% clickbait and 48.4% non-clickbait ratio. With an 80-20 train-test split, four machine learning models were applied. The best-performing model was the ensemble classifier using hard voting, achieving an accuracy of 84.22%, precision of 85.81%, recall of 82%, and an F1-score of 83.86%. These results indicate that the ensemble classifier outperforms single classifiers in identifying clickbait in Indonesian news headlines. The findings suggest that ensemble classifiers are a promising approach for improving clickbait detection in digital news media.

Keywords: *Clickbait Detection, Ensemble, Machine Learning, Classification Models, Indonesian News Headlines*

1. Pendahuluan

Pesatnya kemajuan di era digital, menyebabkan Informasi pada internet terus meningkat. Salah satu informasi yang terus bertambah adalah teks berita. Informasi berita dapat disebarkan dengan cepat melalui berbagai platform online. Menurut hasil survei yang dilakukan oleh Cindy pada Katadata Media Network menunjukkan pembaca media digital sebagai sumber berita utama masih sebanyak 84% pada tahun 2023 dimana persentase ini sudah menurun drastis dibandingkan tahun-tahun sebelumnya [1], namun tetap saja pembaca media digital masih tinggi dibandingkan dengan pembaca media cetak. Kemudahan dalam berbagi informasi ini juga membawa tantangan baru, salah satunya adalah penyebaran *clickbait*. *Clickbait* adalah sebuah teknik untuk membuat sebuah judul menjadi semenarik mungkin sehingga membuat pembaca tertarik untuk meng-klik judul tersebut karena penasaran untuk membacanya, bahkan judul *clickbait* tersebut membuat banyak pembaca menyebarkan berita tanpa membuka isinya [2].

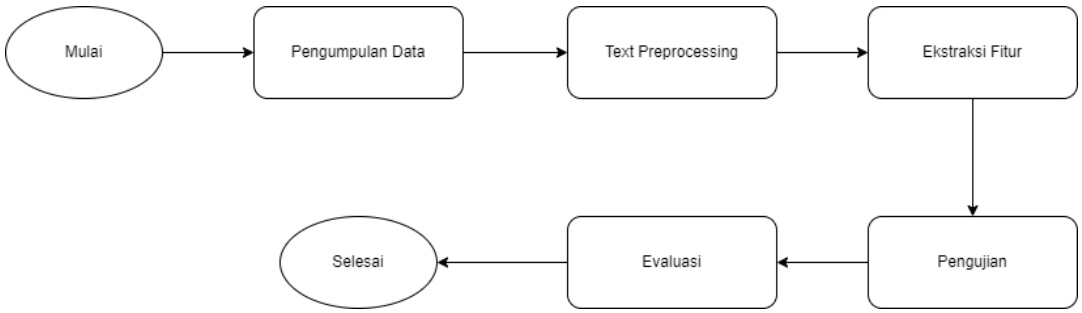
Untuk mengetahui bahwa judul berita tersebut merupakan clickbait, maka diperlukan sebuah alat untuk mengklasifikasi suatu judul tersebut merupakan *clickbait atau tidak. Clickbait detection* dapat menjadi solusi masalah tersebut, sehingga dapat membuat pembaca lebih berhati-hati dalam mengkonsumsi informasi. Penelitian sebelumnya tentang klasifikasi berita clickbait pernah dilakukan oleh Fikri Alwan Ramadhan, dkk dengan judul "Penerapan Metode Multinomial Naïve Bayes untuk Klasifikasi Judul Berita *Clickbait* dengan *Term Frequency - Inverse Document Frequency*". Penelitian ini menggunakan metode Multinomial Naïve Bayes dan TF-IDF (Term Frequency - Inverse Document Frequency) dengan data latih berjumlah 800 data dan data test berjumlah 200 data. Hasil penelitian mendapatkan akurasi sebesar 65%, recall sebesar 65,69% dan precision sebesar 65,69 %[3]. Kemudian penelitian yang dilakukan oleh Shiva Ram Dam, dkk dengan judul "*Detecting Clickbaits on Nepali News using SVM and RF*" menghasilkan F1 score sebesar 0.9483 pada model SVM dan 0.9473 pada model Random Forest [4].

p-ISSN: 2986-3929
e-ISSN: 3032-1948

Berdasarkan paparan penelitian sebelumnya, pada penelitian ini akan membandingkan performa ensemble classifier untuk klasifikasi teks pada berita bahasa Indonesia untuk menentukan *clickbait* atau *non-clickbait* dengan model klasifikasi tunggal. Model klasifikasi tunggal yang akan digunakan dalam penelitian diantaranya Multinomial Naïve Bayes, Support Vector Machine (SVM), dan Random Forest.

2. Metode Penelitian

Penelitian yang dilakukan penulis terdiri dari beberapa tahapan. Pertama adalah pengumpulan data berupa judul berita berbahasa Indonesia dalam bentuk teks. Setelah data atau judul berita berbahasa indonesia telah terkumpul, maka akan melalui tahap text preprocessing. Setelah data telah bersih maka akan masuk ke dalam ekstraksi fitur. Setelah menghasilkan representasi numerik dari data maka akan masuk ke dalam pengujian. Dalam pengujian akan dilakukan uji pada beberapa model yang hasilnya kemudian akan dievaluasi dengan confusion matrix. Secara umum, alur penelitian dapat dilihat pada Gambar 1.

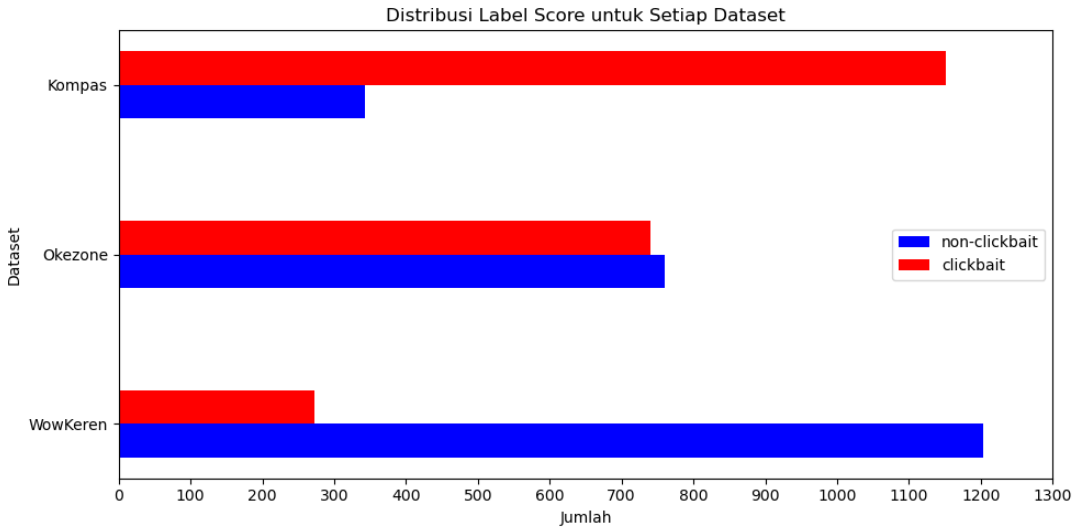


Gambar 1. Alur penelitian

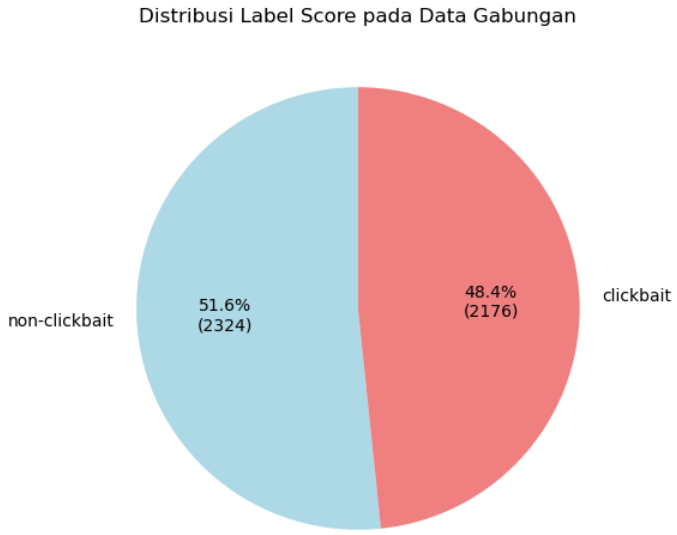
2.1 Pengumpulan Data

Data yang digunakan adalah data yang dapat menunjang penelitian yang berkaitan dengan berita baik yang bersifat *clickbait* atau bukan *clickbait.* Data penelitian ini didapat dari jurnal hasil penelitian oleh Andika William dan Yunita Sari dengan judul "CLICK-ID: *A novel dataset for Indonesian clickbait headlines*" [5]. Dataset CLICK-ID diambil dari 12 penerbit diantaranya adalah detikNews, Fimela, Kapanlagi, Kompas, Liputan6, Republika, Sindonews, Tempo, Tribunnews, Okezone, Wowkeren, and Posmetro-Medan. Dataset CLICK-ID mempunyai 46,517 data yang dibagi menjadi 2 grup yaitu 15,000 yang sudah diberi label (*annotated*) dan 31,517 yang tidak diberi label (*raw*). Pada penelitian ini, akan menggunakan data dari 3 penerbit diantaranya Wowkeren, Okezone, dan Kompas dengan dataset total sebanyak 4500. Atribut pada dataset yang digunakan adalah judul dan label yang bernilai 1 untuk *clickbait* dan 0 untuk *non-clickbait.* Persebaran data tiap penerbit dapat dilihat pada Gambar 2. Dari 4500 data akan dibagi menjadi data clickbait sebanyak 2176 dan data non-clickbait sebanyak 2324 yang dapat dilihat pada Gambar 3.

p-ISSN: 2986-3929
e-ISSN: 3032-1948



Gambar 2. Distribusi label untuk setiap penerbit



Gambar 3. Distribusi label pada data gabungan

2.2 Text Preprocessing

Dalam penelitian ini, data yang didapat akan diolah terlebih dahulu sebelum diuji oleh model. Proses ini bertujuan untuk mengubah teks menjadi bentuk data yang dapat dengan mudah dipahami dan diolah oleh komputer [2]. Dalam *text preprocessing* terdapat beberapa tahapan, tahap pertama yaitu *casefolding* yang bertujuan untuk merubah semua huruf pada teks menjadi huruf kecil (*lowercase*). Tahap ke dua yaitu *Remove Punctuation* yang berfungsi untuk menghilangkan tanda baca pada semua teks.

2.3 Ekstraksi Fitur

Setelah data melakukan text preprocessing, data akan melakukan ekstraksi fitur terlebih dahulu sebelum diuji oleh model. Proses ini bertujuan untuk memberikan bobot pada sebuah fitur yang mencerminkan seberapa penting atau relevan fitur tersebut dalam konteks tertentu [6]. Dengan cara ini, bisa mengukur sejauh mana fitur tersebut berkontribusi dalam analisis data dan

menentukan nilai signifikansinya dalam keseluruhan proses. Metode yang sering digunakan dalam pembobotan kata yaitu TF-IDF [6]. TF-IDF (*Term Frequency - Inverse Document Frequency*) adalah sebuah metode yang mengombinasikan dua pendekatan untuk menetapkan bobot pada sebuah kata, yaitu dengan menghitung frekuensi kemunculan kata tersebut dalam dokumen (TF) dan menghitung kebalikan dari jumlah dokumen yang mengandung kata tersebut (IDF) [6]. Penghitungan TF-IDF dapat dilihat pada persamaan (1) dan (2)

$$TF(d,t) = f(d,t) \quad (1)$$

$$IDF(t) = log\left(\frac{N_d}{df(t)}\right) \quad (2)$$

dimana:

$TF(d,t)$: Term frequency
$f(d,t)$: Frekuensi term t pada dokumen d
$IDF(t)$: Inverse document frequency
N_d : Jumlah dokumen keseluruhan
$df(t)$: Jumlah dokumen yang mengandung term t

Sehingga untuk perhitungan TF-IDF dari suatu kata pada dokumen dapat dilakukan dengan Persamaan (3).

$$TF - IDF = TF(d,t) . IDF(t) \quad (3)$$

2.4 Multinomial Naïve Bayes Classifier

Naïve Bayes adalah algoritma pembelajaran mesin berbasis probabilitas yang digunakan untuk klasifikasi. Prinsip utamanya adalah "naïve" karena mengasumsikan bahwa fitur-fitur dalam dataset independen satu sama lain, artinya kehadiran satu fitur tidak mempengaruhi yang lainnya [6]. Meskipun asumsi ini jarang benar dalam praktik, algoritma Naïve Bayes terbukti efektif dan efisien dalam berbagai aplikasi salah satunya yaitu klasifikasi teks. Salah satu model dari Naïve Bayes yang sering digunakan dalam klasifikasi teks adalah multinomial Naïve Bayes [6].
Multinomial Naïve Bayes adalah model yang dikembangkan dari algoritma Bayes yang ideal untuk klasifikasi teks atau dokumen. Dalam formula Multinomial Naïve Bayes Classifier, kelas dokumen tidak hanya ditentukan oleh kata-kata yang muncul, tetapi juga oleh seberapa sering kata-kata tersebut muncul [7]. Probabilitas suatu dokumen d berada di kelas c dapat dihitung menggunakan Persamaan (4)

$$P(c|d) = P(c) \prod_{k=1}^{n} P(T_k|c) \quad (4)$$

dimana:

$P(c|d)$: Probabilitas dokumen d berada di kelas c
$P(c)$: Prior probability suatu dokumen berada di kelas c
$\{t1, t1, t1, \dots, tn\}$: Token dalam dokumen d yang merupakan bagian dari vocabulary dengan jumlah n
$P(T_k|c)$: Probabilitas bersyarat term T_k berada di dokumen pada kelas c

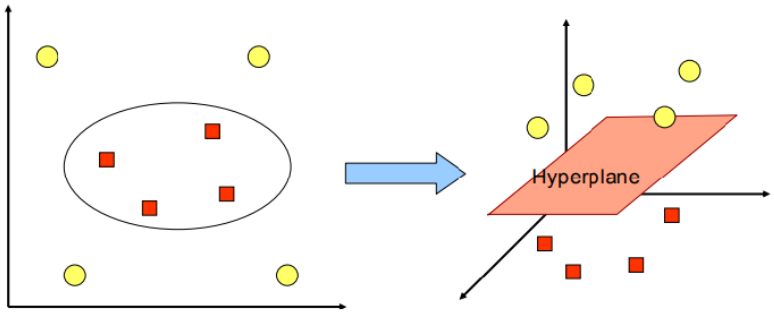
Tujuan klasifikasi dokumen adalah untuk menemukan kelas yang paling sesuai untuk sebuah dokumen. Dalam klasifikasi Naïve Bayes, kelas terbaik ditentukan dengan mencari maximum a posteriori (MAP) untuk kelas yang disebut c_{map} melalui Persamaan (5).

$$c_{map} = \underset{c \in C}{arg\,max}\ \hat{P}(c) \prod_{k=1}^{n} \hat{P}(T_k|c) \quad (5)$$

p-ISSN: 2986-3929
e-ISSN: 3032-1948

2.5 Support Vector Machine

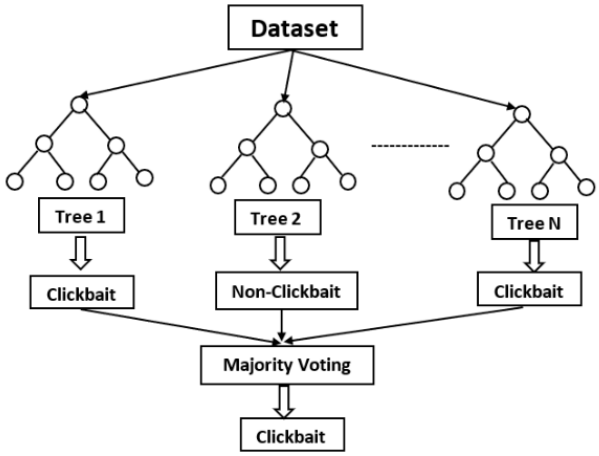
Support Vector Machine (SVM) adalah metode *machine learning* yang bekerja berdasarkan prinsip *Structural Risk Minimization* (SRM) dengan tujuan menemukan *hyperplane* terbaik untuk memisahkan dua kelas pada ruang input. Dalam penelitian ini, metode SVM digunakan untuk pemodelan klasifikasi. Mengelompokkan data atau objek ke dalam kelas atau label berdasarkan fitur-fitur tertentu disebut klasifikasi. Untuk pemodelan klasifikasi, algoritma SVM memiliki konsep yang lebih sistematis dan matang dibandingkan dengan algoritma lain. SVM juga mampu mengatasi masalah klasifikasi baik dengan metode linear maupun non-linear. Dalam algoritma SVM, *hyperplane* berfungsi sebagai pemisah antara satu kelas dan kelas lainnya. *Hyperplane* adalah garis yang memisahkan data antara kelas atau kategori. *Hyperplane* dikatakan ideal jika berada di tengah-tengah dua kelas, yang berarti memiliki jarak terjauh ke data eksternal dari kedua kelas tersebut [8]. *Hyperplane* dapat dilihat pada Gambar 4.



Gambar 4. Hyperplane pada model SVM [9]

2.6 Random Forest

Random Forest (RF) adalah algoritma klasifikasi terarah yang menciptakan hutan dengan berbagai pohon keputusan. Konsep dasar algoritma ini adalah membangun sejumlah pohon keputusan dari set pelatihan yang dipilih secara acak. Algoritma RF membangun banyak pohon keputusan, di mana setiap pohon memberikan suara kemudian semua suara dikumpulkan dan akan dilakukan klasifikasi [4]. Gambaran umum algoritma random forest dapat dilihat pada Gambar 5.



Gambar 5. Struktur algoritma Random Forest [4]

2.7 Ensemble Classifier

Ensemble learning adalah teknik dalam machine learning yang memanfaatkan beberapa model klasifikasi yang berbeda untuk meningkatkan akurasi prediksi. Pada ensemble learning, setiap model klasifikasi dirancang untuk membuat prediksi yang berbeda, lalu hasil-hasil tersebut digabungkan untuk mendapatkan hasil yang lebih akurat. Teknik ini dapat meningkatkan akurasi prediksi dengan mengurangi bias serta meningkatkan robustness model klasifikasi terhadap variasi data [10]. Ada beberapa metode dalam ensemble salah satunya yaitu voting. Voting classifier memiliki dua teknik voting, yaitu hard voting dan soft voting. Pada penelitian ini, peneliti akan menggunakan teknik hard voting classifier. Dalam hard voting, prediksi akhir ditentukan melalui voting mayoritas, di mana hasil akhir memilih prediksi kelas yang paling sering muncul di antara model-model dasar [11]. Pada panelitian ini, model klasifikasi dasar yang akan digunakan yaitu Multinomial Naïve Bayes, Support Vector Machine, dan Random Forest

2.8 Confusion Matrix

Confusion matrix menyajikan informasi tentang nilai aktual dan prediksi yang dihasilkan oleh model. Matriks ini menyajikan data terkait nilai aktual dan nilai prediksi dari hasil klasifikasi berita yang didapatkan dari proses pengklasifikasian. Ukuran utama yang digunakan adalah akurasi klasifikasi, yang dihitung sebagai jumlah kasus yang diklasifikasikan dengan benar pada set pengujian dibagi dengan jumlah total kasus dalam set pengujian. Precision dan recall mengukur sejauh mana klasifikasi ini tepat dan seberapa lengkap klasifikasi ini pada kelas yang positif [7]. *Accuracy*, *recall*, *precision, dan F1-Score* dapat dihitung menggunakan rumus sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

dimana:

TP (True Positive) : jumlah prediksi yang tepat bahwa kelas bersifat positif
TN (True Negative) : jumlah prediksi yang tepat bahwa kelas bersifat negatif
FP (False Positive) : jumlah prediksi yang salah bahwa kelas bersifat positif
FN (False Negative) : jumlah prediksi yang salah bahwa kelas bersifat negatif

3. Hasil dan Diskusi

Setelah melakukan tahap text preprocessing dan ekstraksi fitur pada dataset, langkah selanjutnya adalah membagi dataset menjadi data training dan testing masing-masing sebesar 80% dan 20%. Pada tahap ini, data training digunakan untuk melatih model, sementara data testing digunakan untuk menguji kinerja model yang sudah dilatih. Penelitian ini menggunakan tiga model tunggal yaitu Multinomial Naïve Bayes, Support Vector Machine, dan Random Forest. Selain itu, juga dilakukan pengujian dengan pendekatan ensemble untuk melihat apakah gabungan dari beberapa model dapat memberikan hasil yang lebih baik.

3.1 Pengumpulan Data

Pada penelitian akan digunakan data annotated pada dataset CLICKID. Adapun contoh data dapat dilihat pada Tabel 1.

p-ISSN: 2986-3929
e-ISSN: 3032-1948

Tabel 1. Contoh data annotated

Atribut	Data 1	Data 2
title	Diduga Pasok Sabu di Anambas, Adik Wali Kota Tanjungpinang Ditangkap	Penemuan Terbaru Ungkap Penyebab Punahnya Dinosaurus
label	non-clickbait	clickbait
label_score	0	1

3.2 Text Preprocessing

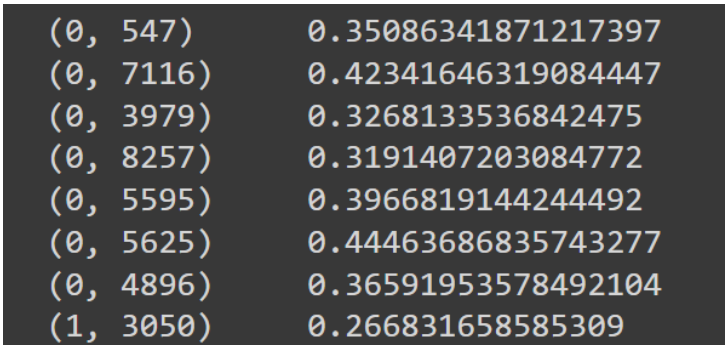
Setelah mendapatkan data, selanjutnya akan dilakukan preprocessing untuk menghilangkan kolom pada dataset yang tidak digunakan serta merubah teks menjadi lowercase dan menghilangkan tanda baca. Hasil dari data preprocessing dapat dilihat pada Tabel 2.

Tabel 2. Hasil data preprocessing

title	label_score
diduga pasok sabu di anambas adik wali kota tanjungpinang ditangkap	0
penemuan terbaru ungkap penyebab punahnya dinosaurus	1

3.3 Ekstraksi Fitur

Teknik ekstraksi fitur menggunakan TF-IDF dilakukan untuk mengubah data tekstual menjadi representasi numerikal sebagai input dari tiap model. Hasil dari ekstraksi fitur dapat dilihat pada Gambar 6.

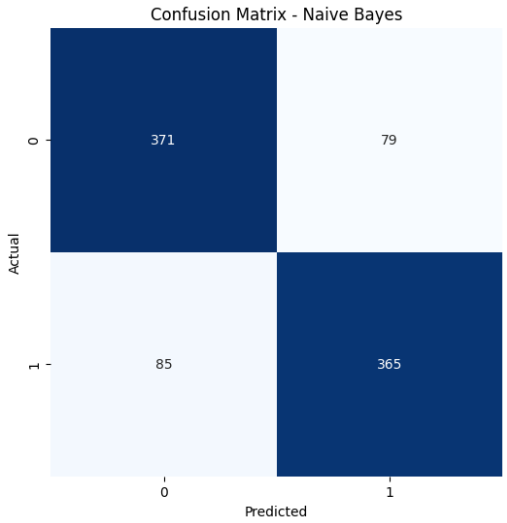


Gambar 6. Hasil ekstraksi fitur dengan TF-IDF

3.4 Hasil model Multinomial Naïve Bayes

Pada percobaan pertama dengan model Multinomial Naïve Bayes (MNB), hasil yang diperoleh dari pengujian model dapat dilihat dalam bentuk confusion matrix yang dapat dilihat pada Gambar 7.

p-ISSN: 2986-3929
e-ISSN: 3032-1948



Gambar 7. Confusion matrix model Multinomial Naïve Bayes

Berdasarkan confusion matrix pada Gambar 7, peneliti bisa mengidentifikasi komponen True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Komponen-komponen ini kemudian digunakan untuk mengevaluasi model dengan metrik seperti *accuracy*, *recall*, *precision*, dan *F1-score* yang dapat dilihat pada Tabel 3.

Tabel 3. Evaluasi model Multinomial Naïve Bayes

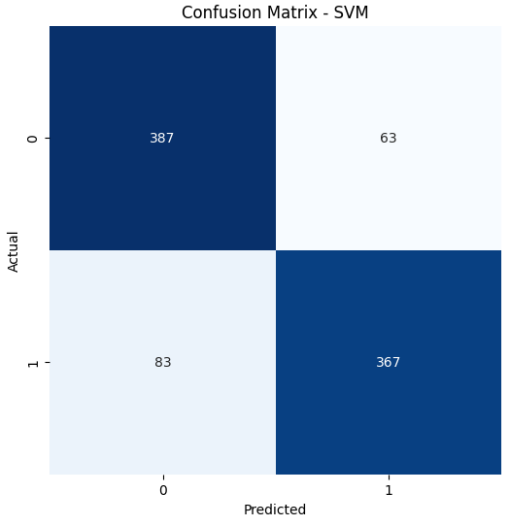
Accuracy	Recall	Precision	F1-Score
81,78%	81,11%	82,21%	81,66%

Dari hasil evaluasi pada Tabel 3, dapat dilihat bahwa model Multinomial Naïve Bayes sudah cukup baik dalam melakukan deteksi *clickbait*.

3.5 Hasil model Support Vector Machine

Model selanjutnya yang digunakan yaitu Support Vector Machine (SVM), hasil confusion matrix yang didapat dapat dilihat pada Gambar 8.

p-ISSN: 2986-3929
e-ISSN: 3032-1948



Gambar 8. Confusion matrix model SVM

Berdasarkan confusion matrix yang dihasilkan oleh model SVM yang ditampilkan pada Gambar 8, kemudian dapat dihitung metrik *accuracy*, *recall*, *precision*, dan *F1-score* sebagai bagian dari evaluasi model. Hasil evaluasi ini dapat dilihat pada Tabel 4.

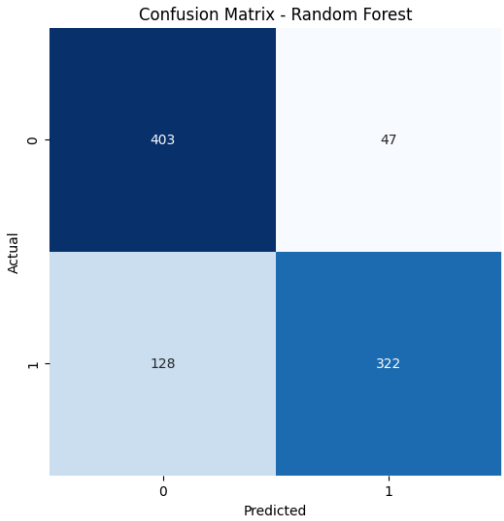
Tabel 4. Evaluasi model SVM

Accuracy	Recall	Precision	F1-Score
83,78%	81,56%	85,35%	83,41%

Dari hasil evaluasi pada Tabel 4, dapat dilihat bahwa model SVM sedikit lebih unggul dibandingkan Multinomial Naïve Bayes dalam melakukan deteksi *clickbait*.

3.6 Hasil model Random Forest

Model selanjutnya yang digunakan yaitu Random Forest (RF), confusion matrix yang dihasilkan dari pengujian dapat dilihat pada Gambar 9.



Gambar 9. Confusion matrix model RF

p-ISSN: 2986-3929
e-ISSN: 3032-1948

Pelatihan yang menggunakan model Random Forest menghasilkan confusion matrix seperti pada Gambar 9. Dari confusion matrix ini, kemudian dilakukan perhitungan untuk mendapatkan komponen metrik evaluasi seperti *accuracy*, *recall*, *precision*, dan *F1-score*. Hasil dari perhitungan keempat metrik ini dapat ditemukan pada Tabel 5.

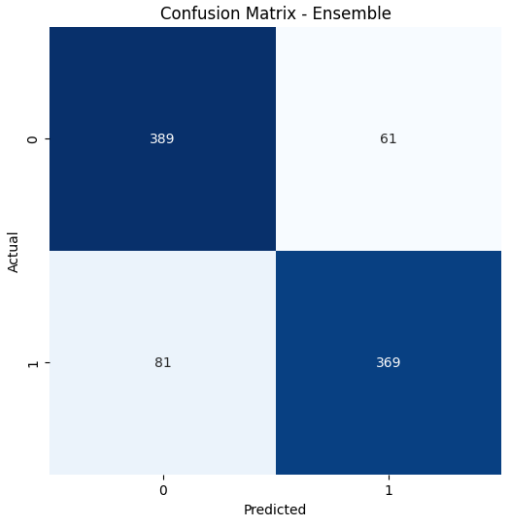
Tabel 5. Evaluasi model RF

Accuracy	Recall	Precision	F1-Score
80,56%	71,56%	87,26%	78,63%

Dari hasil evaluasi pada Tabel 5, dapat dilihat bahwa nilai *precision* pada model Random Forest lebih unggul dari pada dua model sebelumnya. Namun nilai *accuracy*, *recall*, dan *F1-score* masih rendah dibandingkan dengan dua model sebelumnya.

3.7 Hasil ensemble classifier

Pengujian terakhir pada penelitian ini dilakukan dengan menggunakan ensemble classifier dengan metode hard voting yang mengombinasikan tiga model sebelumnya, yaitu Multinomial Naïve Bayes, Support Vector Machine (SVM), dan Random Forest. Hasil confusion matrix yang diperoleh dari pengujian dapat dilihat pada Gambar 10.



Gambar 10. Confusion matrix ensemble classifier

Berdasarkan confusion matrix yang dihasilkan oleh ensemble dengan metode hard voting yang ditampilkan pada Gambar 10, kemudian dapat dihitung metrik accuracy, recall, precision, dan F1-score sebagai bagian dari evaluasi. Hasil evaluasi ini dapat dilihat pada Tabel 6.

Tabel 6. Evaluasi ensemble classifier

Accuracy	Recall	Precision	F1-Score
84,22%	82%	85,81%	83,86%

Dari hasil evaluasi pada Tabel 6, dapat dilihat bahwa nilai accuracy, recall, dan F1-score mendapatkan hasil lebih baik dari ketiga model sebelumnya. Nilai precision masih lebih kecil dibandingkan dengan model Random Forest, namun nilai ini lebih tinggi dari model Multinomial Naïve Bayes dan Support Vector Machine.

3.8 Perbandingan semua model

Setelah melakukan semua pengujian, selanjutnya akan membandingkan semua hasil evaluasi didasarkan pada empat metrik evaluasi yaitu accuracy, precision, recall dan F1-score. Perbandingan evaluasi dapat dilihat pada Tabel 7.

Tabel 7. Hasil Evaluasi Semua Metrik

Classifier	Accuracy	Recall	Precision	F1-Score
MNB	81,78%	81,11%	82,21%	81,66%
SVM	83,78%	81,56%	85,35%	83,41%
RF	80,56%	71,56%	**87,26%**	78,63%
Ensemble	**84,22%**	**82%**	85,81%	**83,86%**

Jika dilihat pada Tabel 7, dapat diketahui bahwa semua model memiliki performa yang seimbang. Akan tetapi metrik evaluasi yang dihasilkan oleh model ensemble adalah metrik dengan keseluruhan nilai terbaik meskipun metrik precision lebih rendah sebesar 1,45% dibandingkan model Random Forest. Akan tetapi jika melihat metrik lainnya yaitu accuracy, precision, dan F1-Score, ensemble mendapatkan nilai tertinggi pada ketiga metrik tersebut.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan dan hasil yang diperoleh selama melakukan penelitian dengan dataset yang berisi judul *clickbait* dan *non-clickbait* yang berjumlah 4500 data, dengan perbandingan kelas data berurut sebesar 51,6% dan 48,4% serta pembagian 80% data *training* dan 20% data *testing*, peneliti berhasil mengaplikasikan 4 model *machine learning* untuk mengklasifikasikan judul berita berbahasa indonesia termasuk *clickbait* atau tidak. Keempat model yang diuji adalah Multinomial Naïve Bayes, Support Vector Machine, Random Forest, dan Ensemble Classifier dengan metode hard voting. Setelah melakukan evaluasi terhadap beberapa parameter dari masing-masing model, maka diperoleh model terbaik yaitu ensemble cassifier. Model ini memiliki metrik *accuracy* 84,22%, *precision* 85,81%, *recall* 82%, dan *F1-Score* 83,86%. Keempat metrik yang dihasilkan oleh model ini merupakan metrik terbaik secara keseluruhan dibandingkan dengan model yang lainnya.

Daftar Pustaka

[1] Cindy Mutia Annur, “Meski Trennya Turun, Media Online Tetap Jadi Sumber Berita Utama Masyarakat Indonesia.” Accessed: May 08, 2024. [Online]. Available: https://databoks.katadata.co.id/datapublish/2023/06/16/meski-trennya-turun-media-online-tetap-jadi-sumber-berita-utama-masyarakat-indonesia

[2] I. Nurindar Wardani, M. Ningsih, and R. Zusyana D., “Penggunaan Clickbait Headline Pada Portal Berita Tribunnews.Com,” *Jurnal Komunikasi dan Sosial Humaniora*, vol. 2, no. 1, 2021.

[3] F. A. Ramadhan, S. H. Sitorus, and T. Rismawan, “Penerapan Metode Multinomial Naïve Bayes untuk Klasifikasi Judul Berita Clickbait dengan Term Frequency - Inverse Document Frequency,” *Jurnal Sistem dan Teknologi Informasi (JustIN)*, vol. 11, no. 1, p. 70, Jan. 2023, doi: 10.26418/justin. v11i1.57452.

[4] S. R. Dam, S. P. Panday, and T. B. Thapa, “Detecting Clickbaits on Nepali News using SVM and RF,” in *9th IOE Graduate Conference*, 2021. [Online]. Available: https://www.dcnepal.com/

[5] A. William and Y. Sari, “CLICK-ID: A novel dataset for Indonesian clickbait headlines,” *Data Brief*, vol. 32, p. 106231, Oct. 2020, doi: 10.1016/J.DIB.2020.106231.

[6] A. Sabrani, I. G. W. Wedashwara W., and F. Bimantoro, “Multinomial Naïve Bayes untuk Klasifikasi Artikel Online tentang Gempa di Indonesia,” *Jurnal Teknologi Informasi, Komputer, dan Aplikasinya (JTIKA)*, vol. 2, no. 1, pp. 89–100, Mar. 2020, doi: 10.29303/JTIKA.V2I1.87.

[7] D. H. Kalokasari, I. M. Shofi, and A. H. Setyaningrum, “Implementasi Algoritma Multinomial Naive Bayes Classifier Pada Sistem Klasifikasi Surat Keluar (Studi Kasus: Diskominfo Kabupaten Tangerang),” *Jurnal Teknik Informatika*, vol. 10, no. 2, Oct. 2017, doi: 10.15408/JTI.V10I2.6199.

[8] A. Ahdika and N. Artiyati, “Implementation of the Support Vector Machine (SVM) Method for Classification of Underdeveloped Areas in Indonesia,” 2024, doi: 10.30865/mib. v8i2.7598.

[9] V. A. N. Syafika and R. D. L. N. Karisma, “Implementasi Support Vector Machine (SVM) dalam Penentuan Klasifikasi Indeks Khusus Penanganan Stunting di Indonesia,” *Seminar Nasional Official Statistics*, vol. 2023, no. 1, 2023, doi: 10.34123/semnasoffstat. v2023i1.1595.

[10] D. S. Sisodia, “Ensemble learning approach for clickbait detection using article headline features,” *Inf Sci*, vol. 22, no. 2019, pp. 31–44, 2019, doi: 10.28945/4279.

[11] S. Kumari, D. Kumar, and M. Mittal, “An ensemble approach for classification and prediction of diabetes mellitus using soft voting classifier,” *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 40–46, Jun. 2021, doi: 10.1016/J.IJCCE.2021.01.001.